Self-learning Adaptive Optics Control

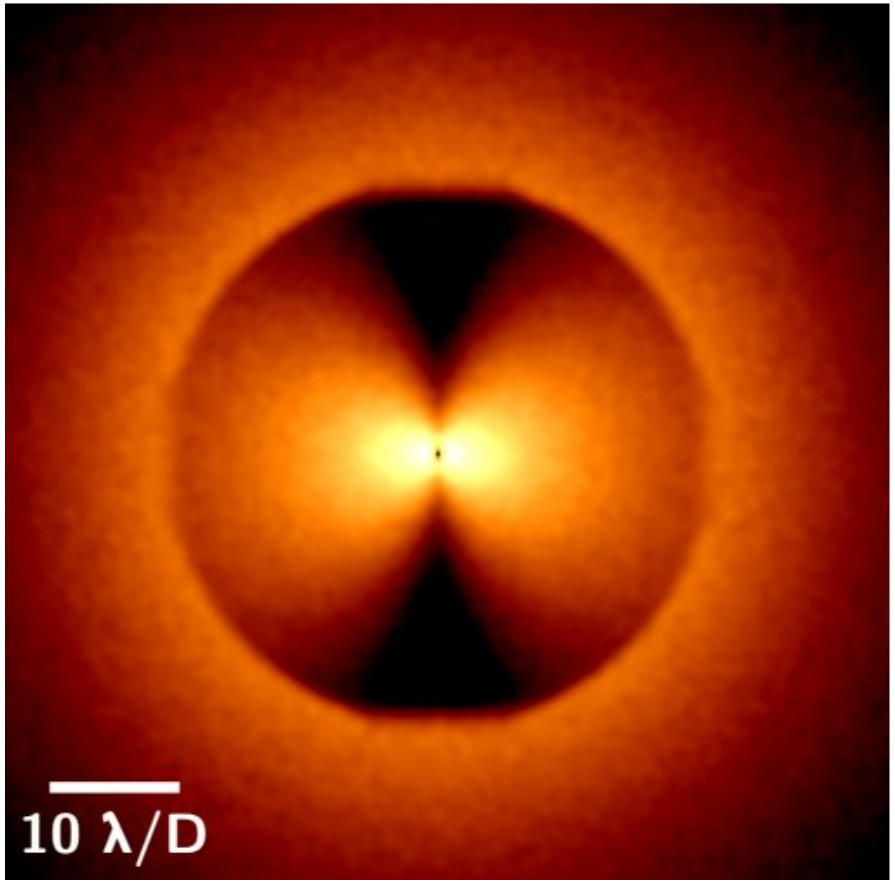
Rico Landman (rlandman@strw.leidenuniv.nl)
Sebastiaan Haffert, Vikram Radhakrishnan, Christoph Keller

14/10/2020

Universiteit Leiden
Leiden Observatory

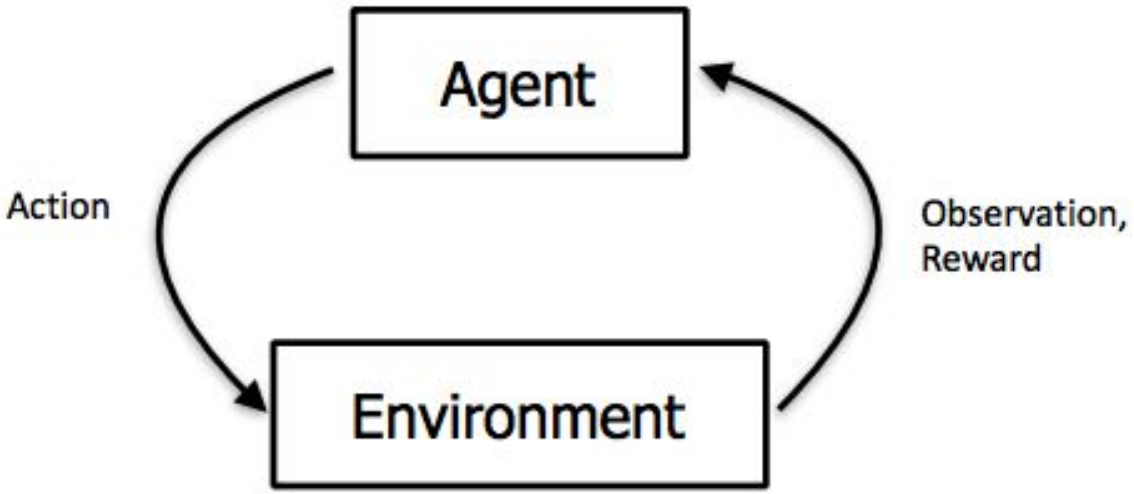
Temporal error

- Deformable mirror lags behind changing atmosphere
- Wind driven halo
- Telescope vibrations
- Predictive control needed



[Cantalloube et al. 2018]

Reinforcement Learning

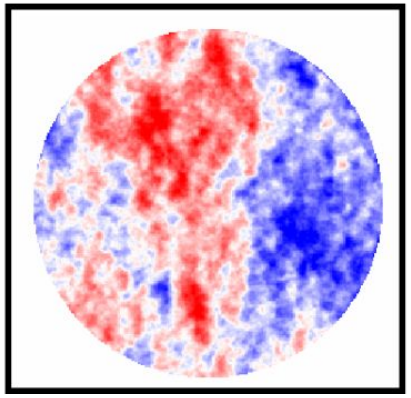


Agent: Deformable Mirror controller

Environment: Atmosphere + system dynamics

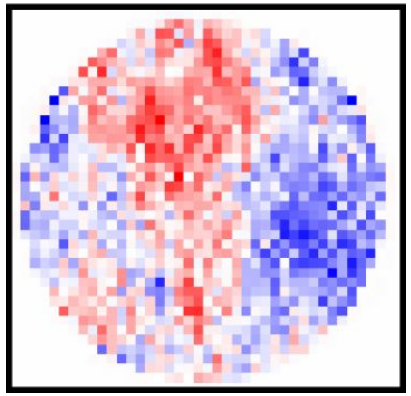
Observation

Reconstructed wavefront

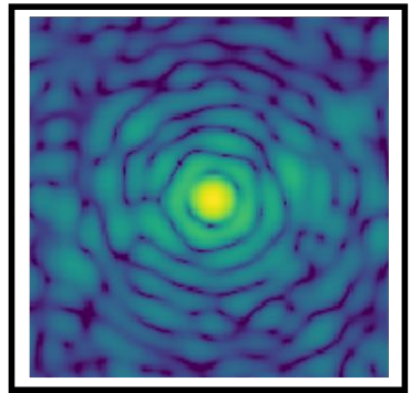


Action

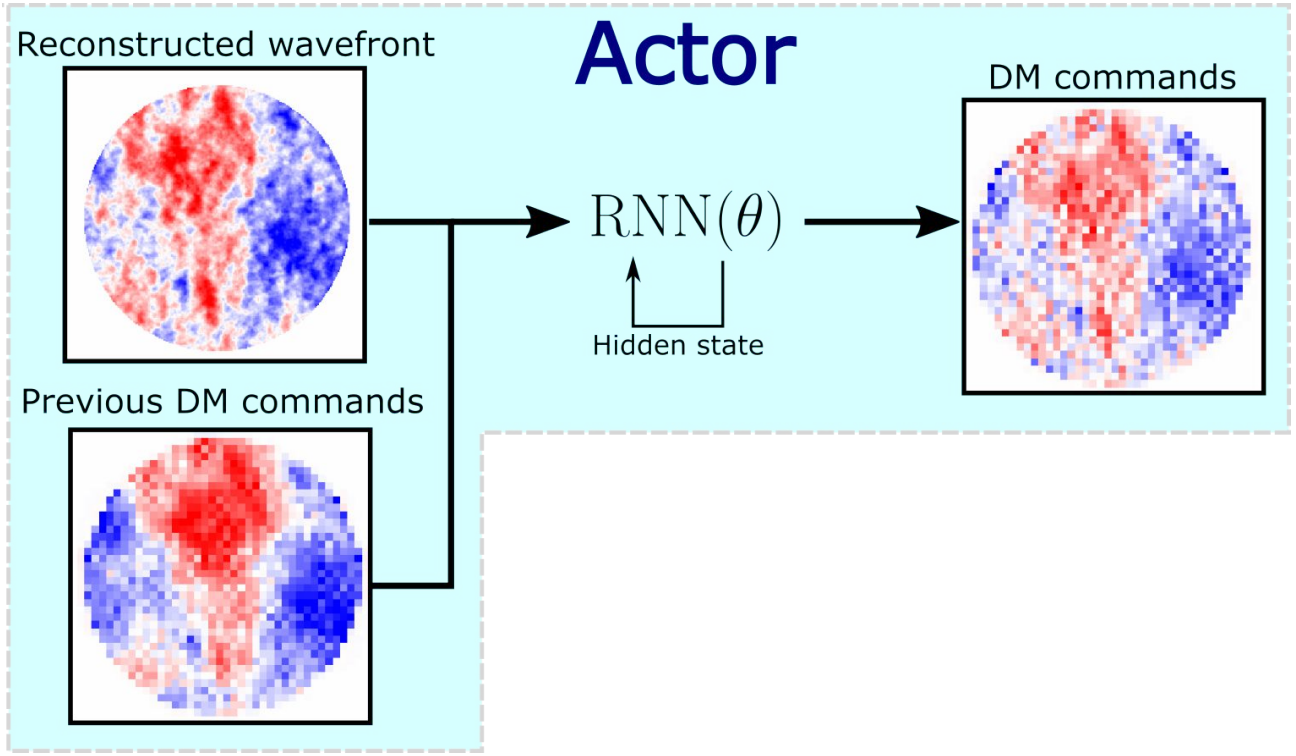
DM commands



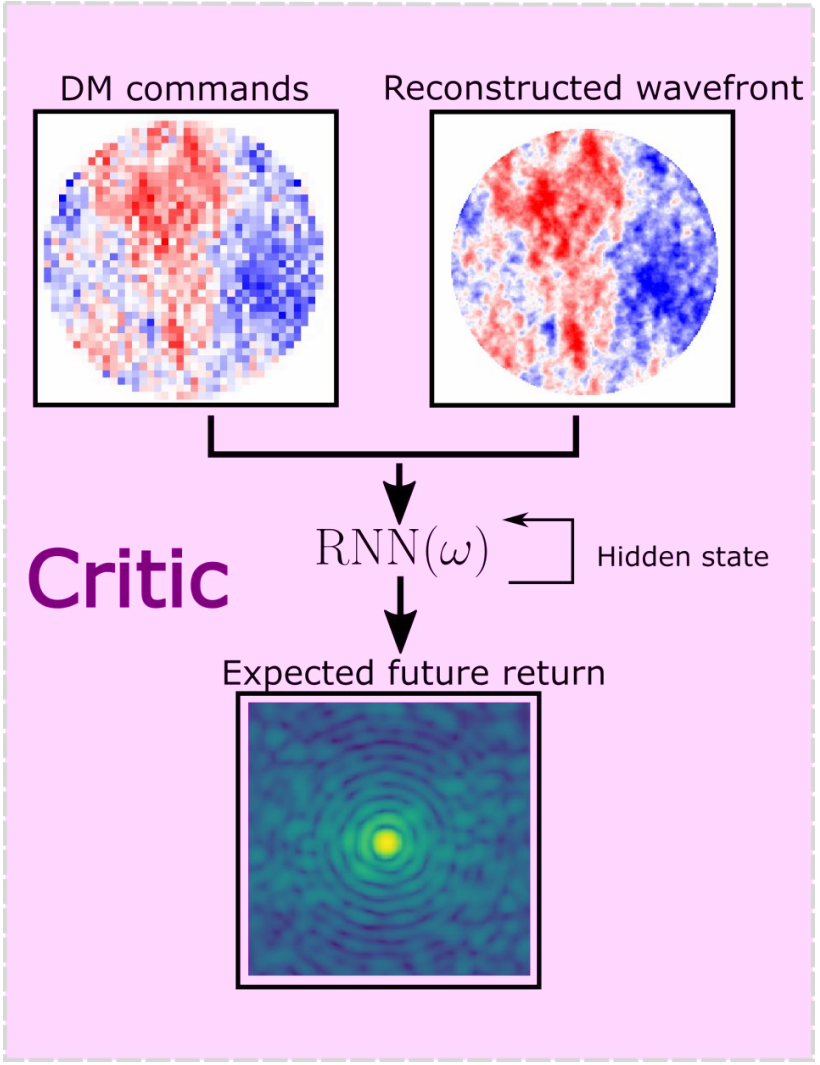
Reward



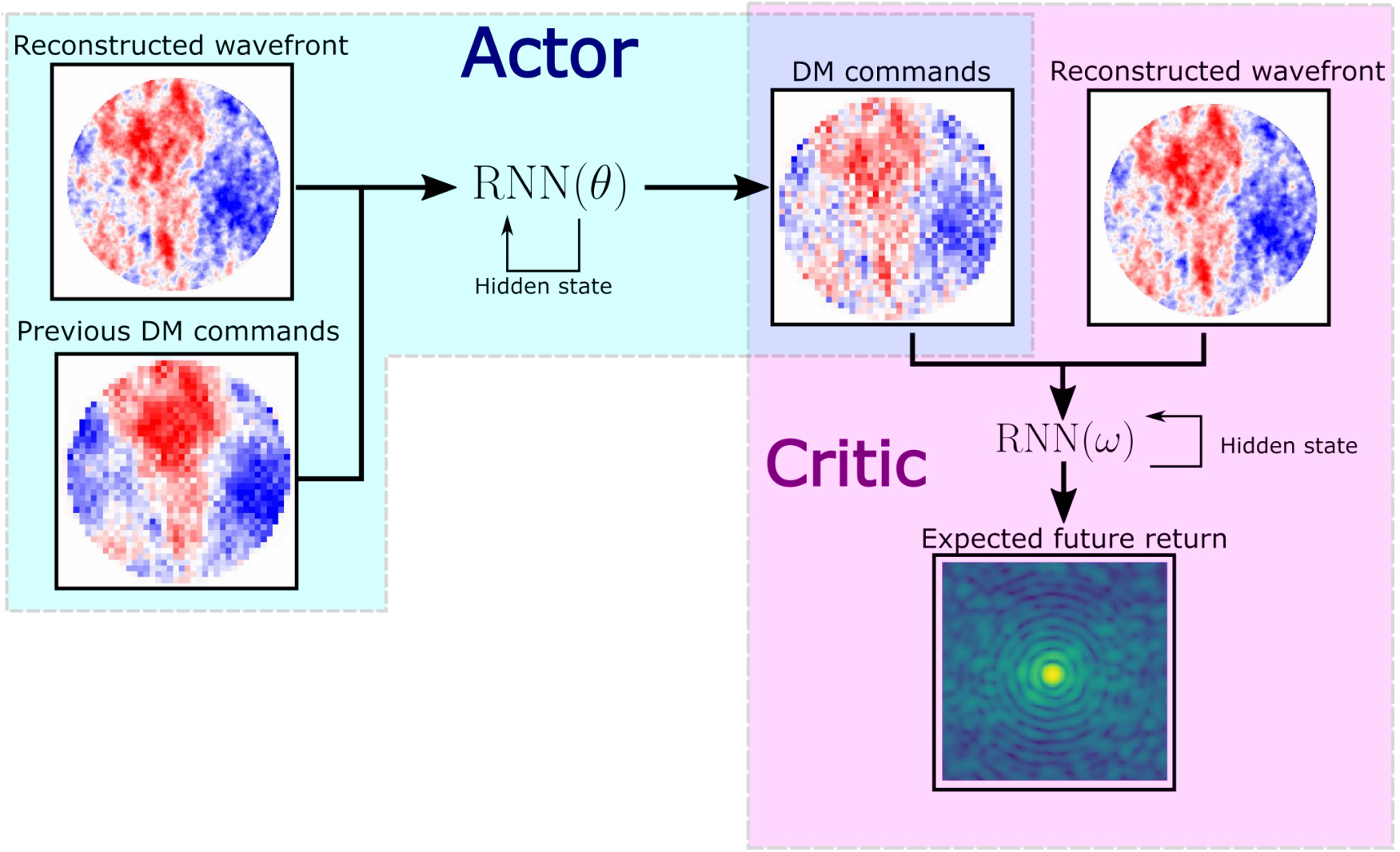
Deterministic Policy Gradient



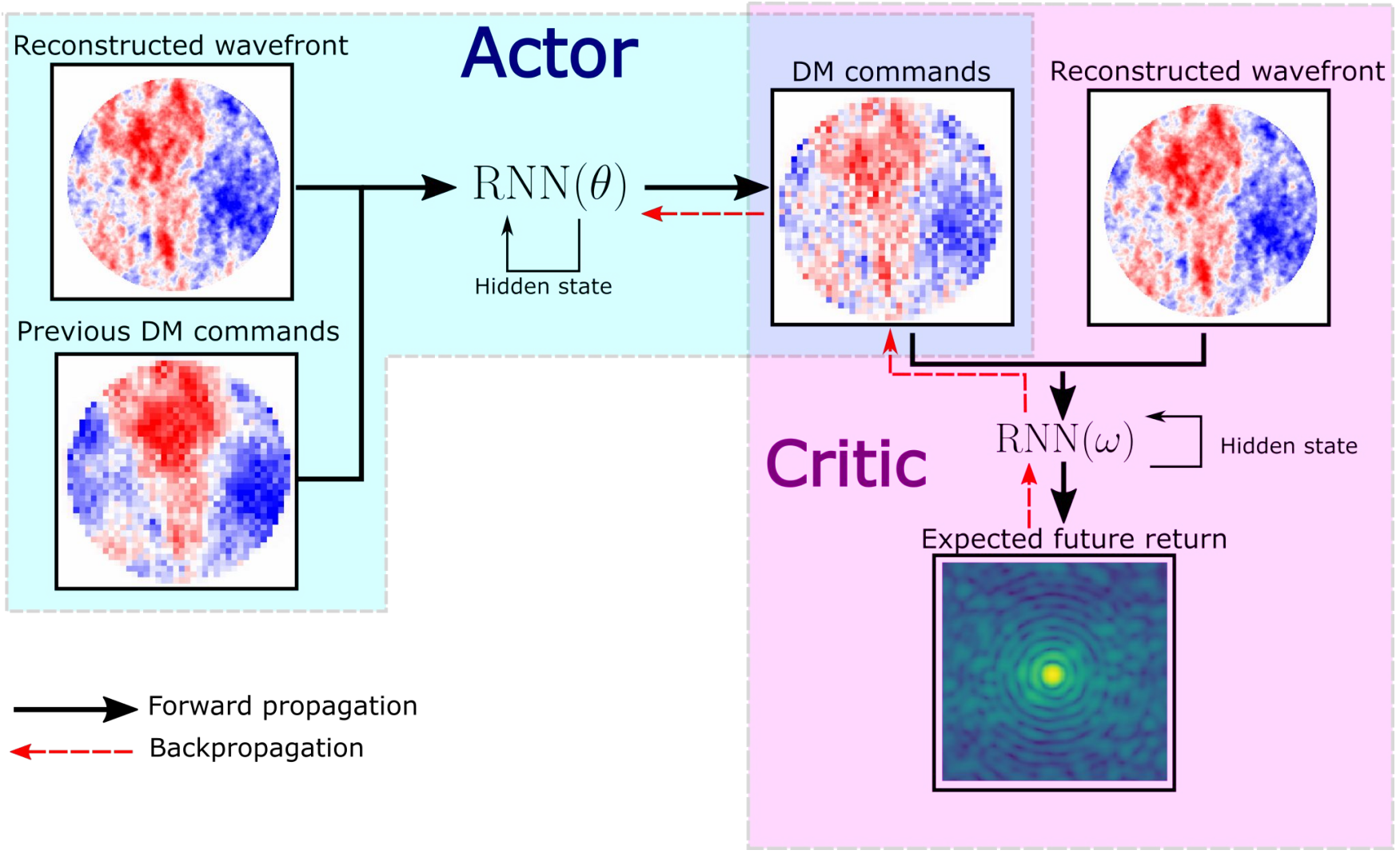
Deterministic Policy Gradient



Deterministic Policy Gradient



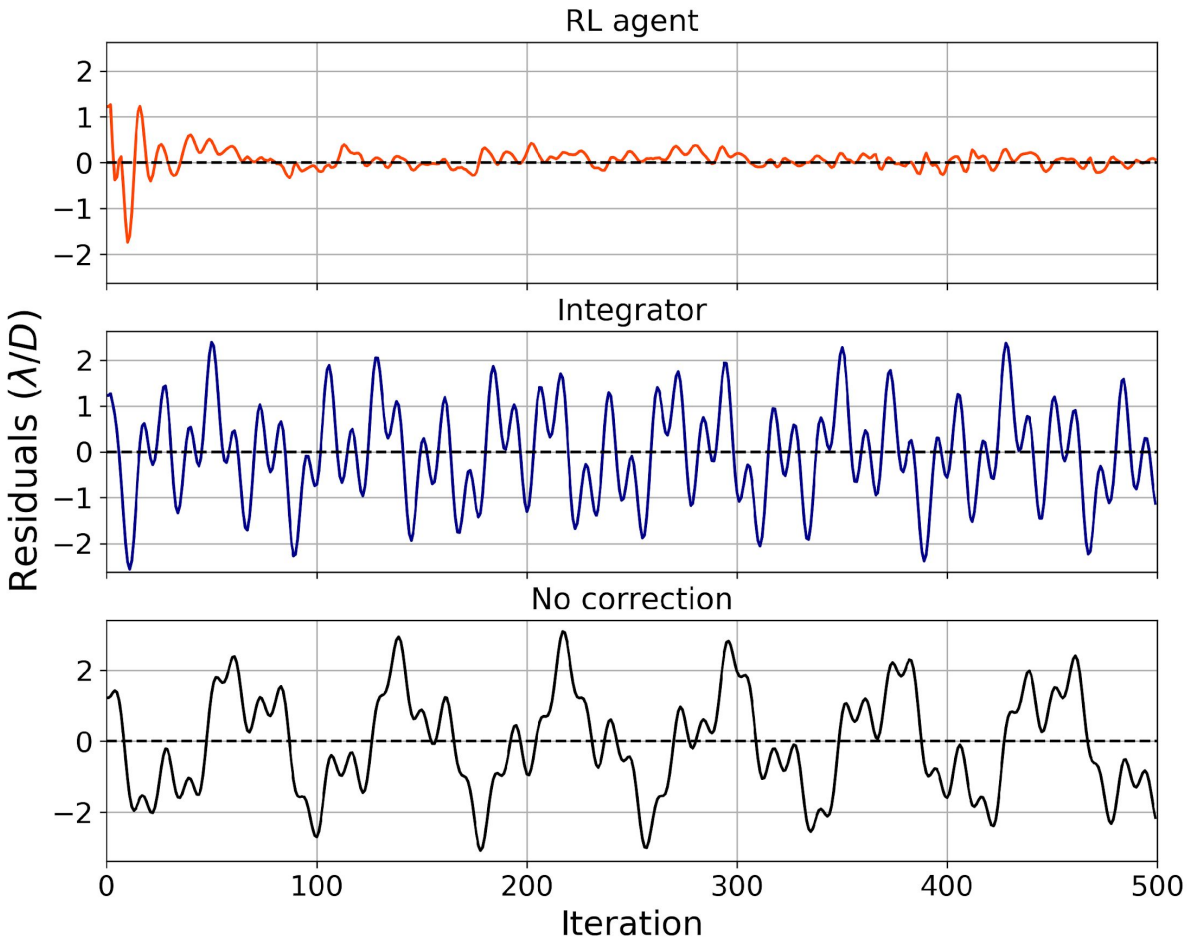
Deterministic Policy Gradient



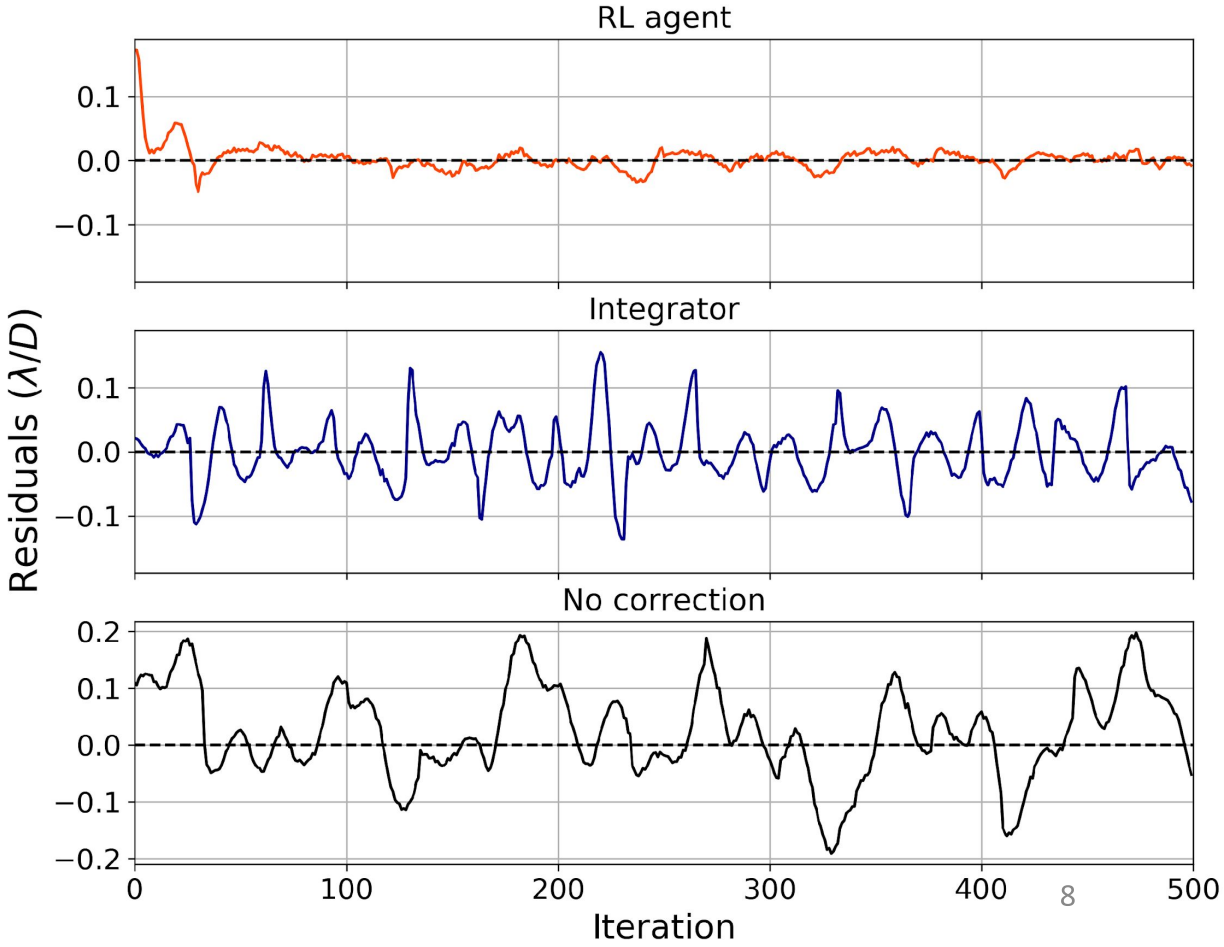
Suppression of tip-tilt vibrations

- Optimization objective: PSF center deviation

Simulations

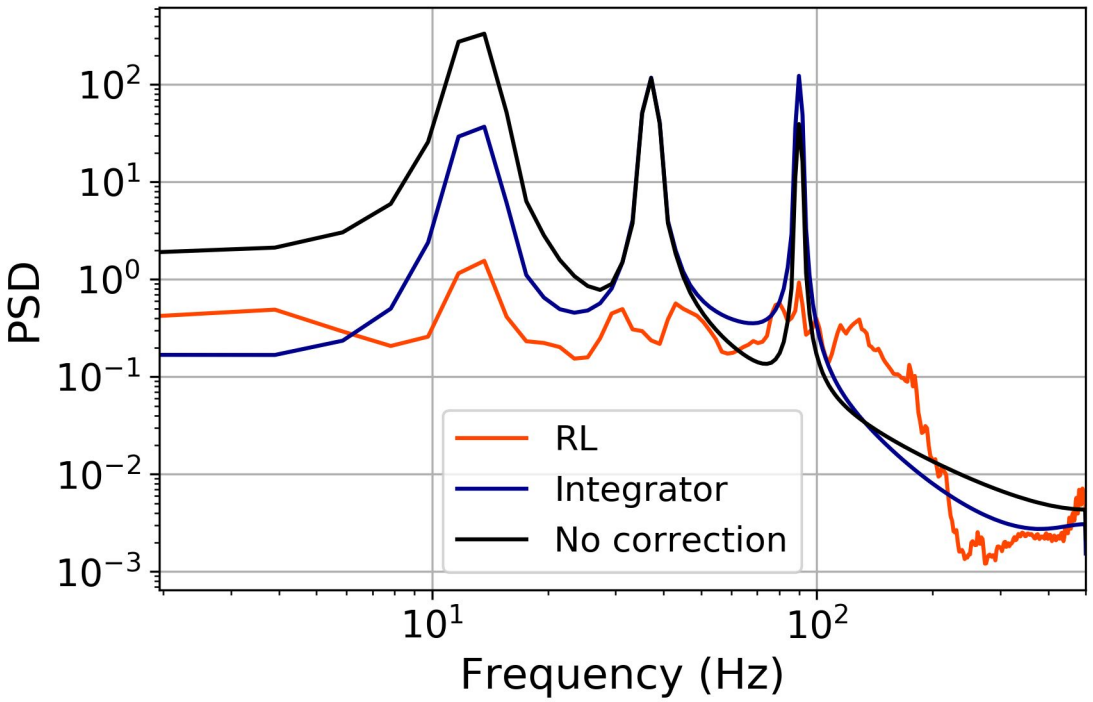


Lab

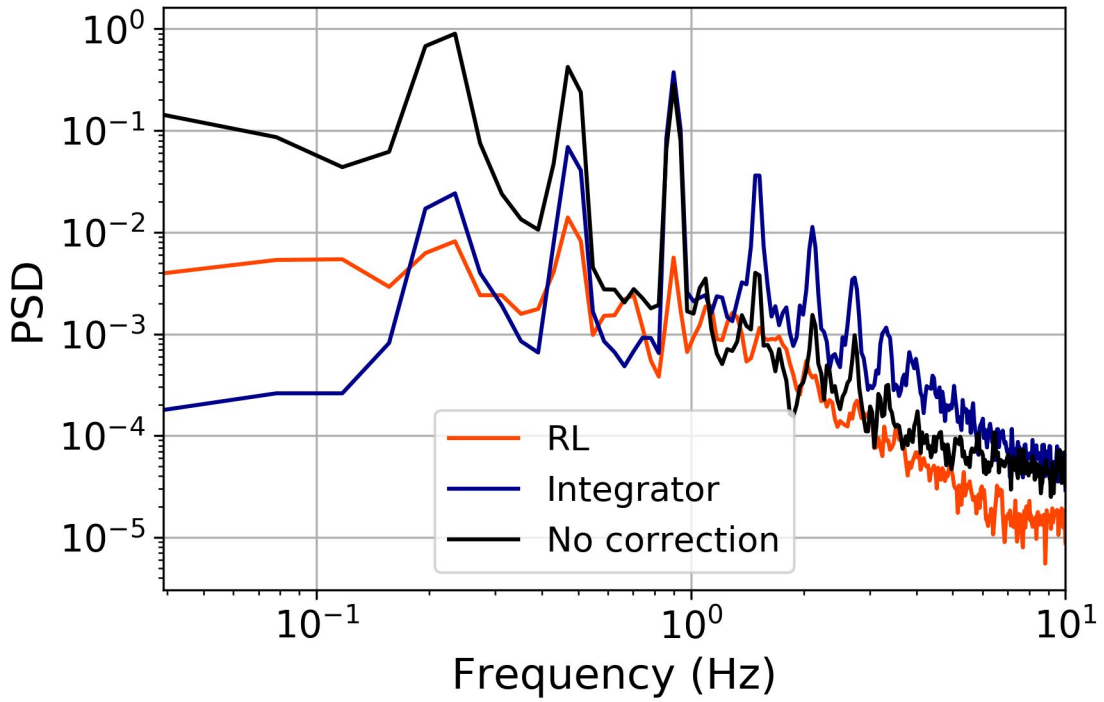


Suppression of tip-tilt vibrations

Simulations

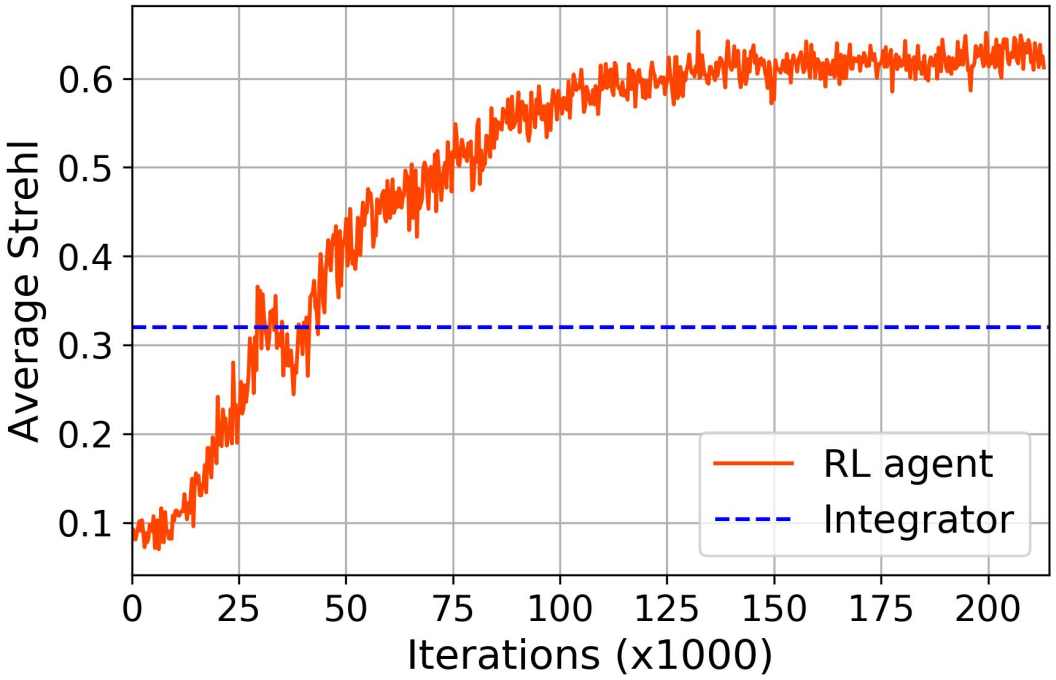


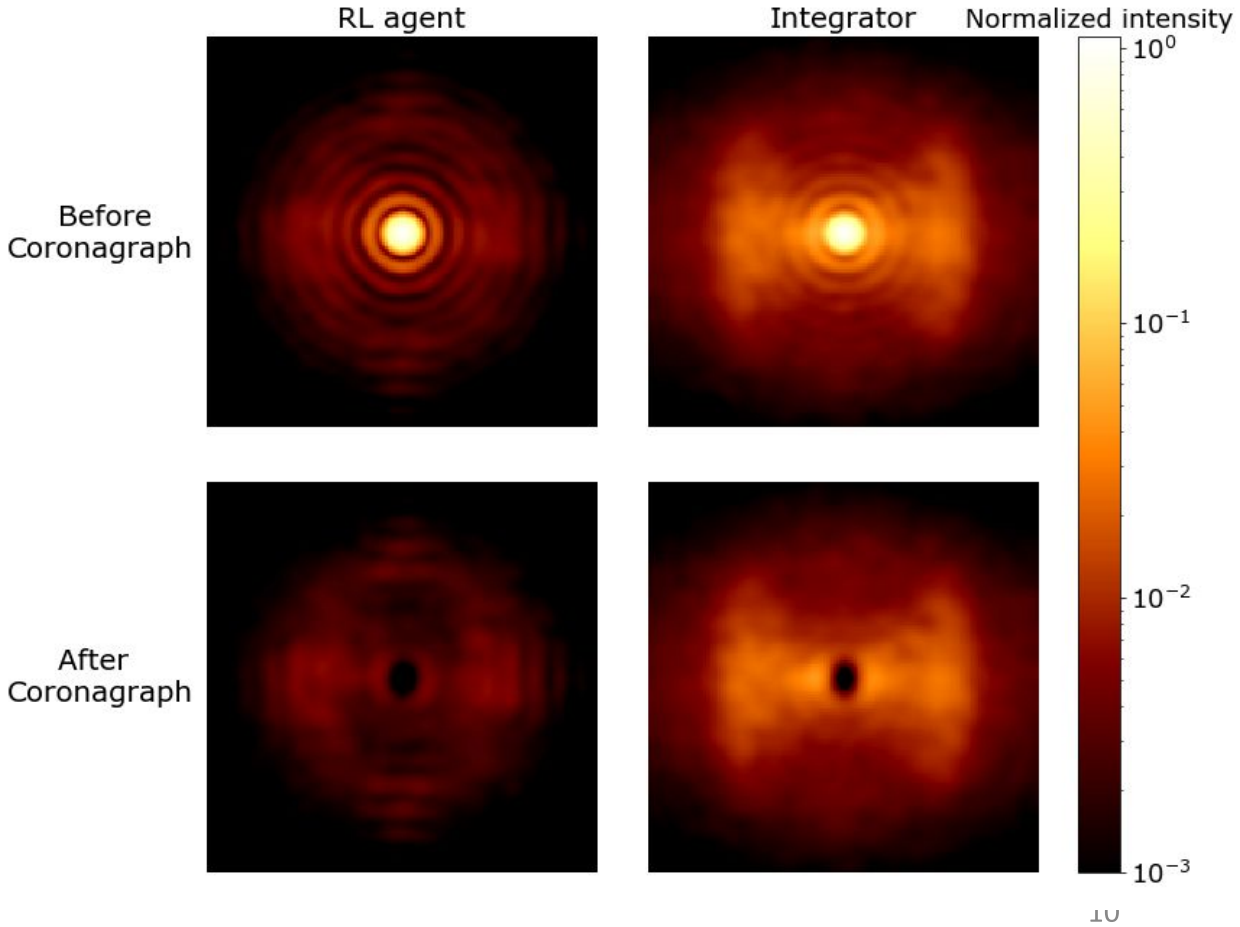
Lab



Simulations: Full wavefront control

- Exaggerated wind speed
- Optimization objective: Strehl ratio





Conclusions

- Reinforcement Learning control can be used to:
  - Suppress tip-tilt vibrations.
  - Reduce the wind driven halo.

- Key properties:
  - **Nonlinear** (PyWFS)
  - **Flexible** (Free to choose reward and inputs)
  - **Model-free** (but allows for incorporation of prior knowledge)
  - Relatively **efficient real-time control**
- But: more research on performance under realistic conditions and practicality is needed.

Extra slides

Actor-Critic

- Actor-critic:
  - **Actor**: Maps observations to actions (Controller)

$$\text{Observation} \xrightarrow{\theta} \text{Action}$$

  - **Critic**: Estimates value of action for given observation (Cost function)

$$\text{Observation, Action} \xrightarrow{\omega} \text{Value}$$

- Optimize Actor parameters with gradients obtained from Critic:

$$\frac{d\text{Value}}{d\theta} = \frac{d\text{Value}}{d\text{Action}} \frac{d\text{Action}}{d\theta}$$

- Actor and critic may be any differentiable parameterized model (e.g. Neural Networks).

Predictive control and closed-loop AO

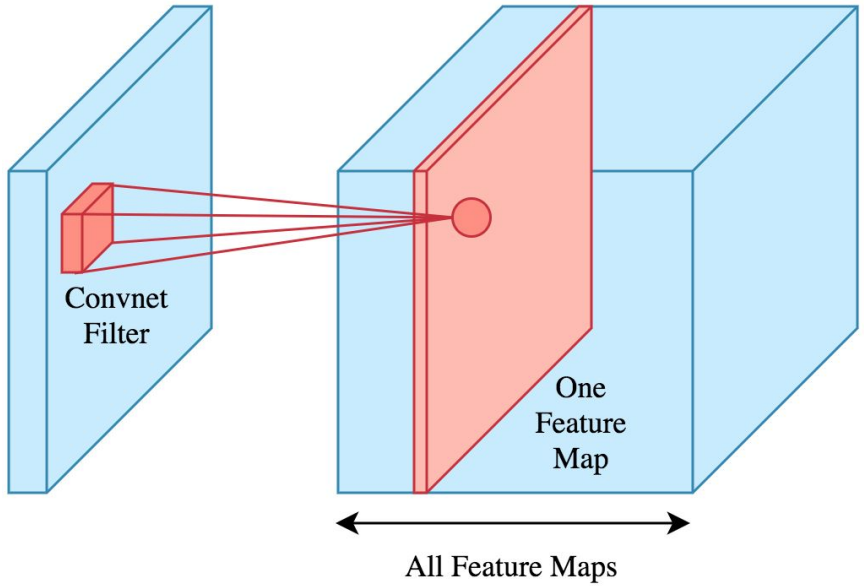
- Only observe closed-loop residuals
  - DM or atmosphere?

- Often used: pseudo open-loop data.
  → BUT: - Requires knowledge of servo-lag etc.
  - 1% error in open-loop wavefront still significant to closed-loop.
  - Not trivial for nonlinear WFS (Pyramid)

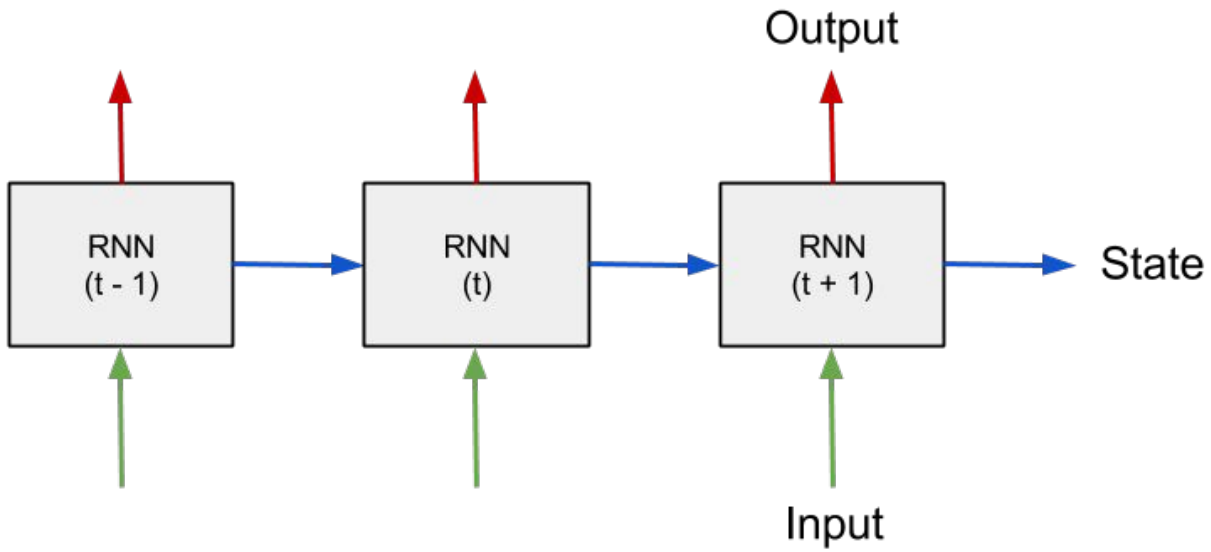
- Partially Observable Markov Decision Process.
  - Recurrent Neural Networks.

Intermezzo: Neural Networks

Spatial data:
Convolutional Neural Networks
(CNN)



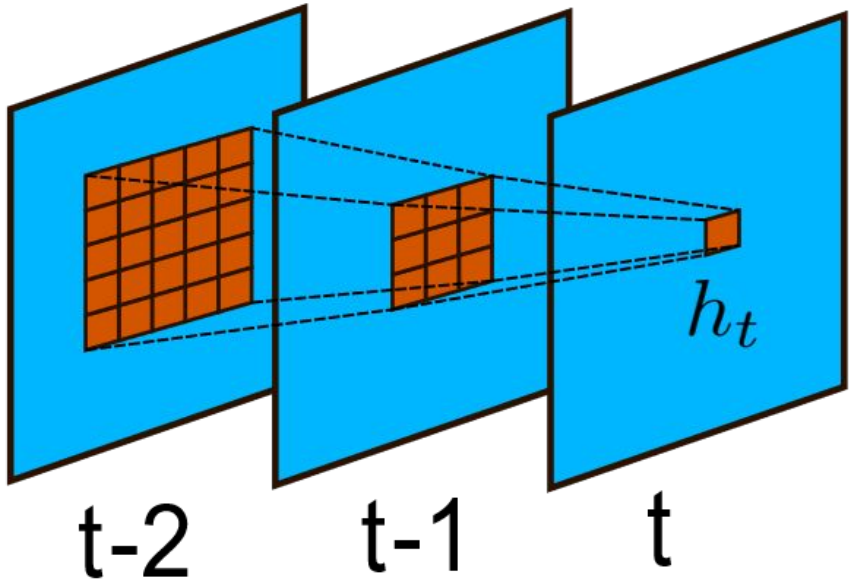
Temporal data:
Recurrent Neural Networks
(RNN)



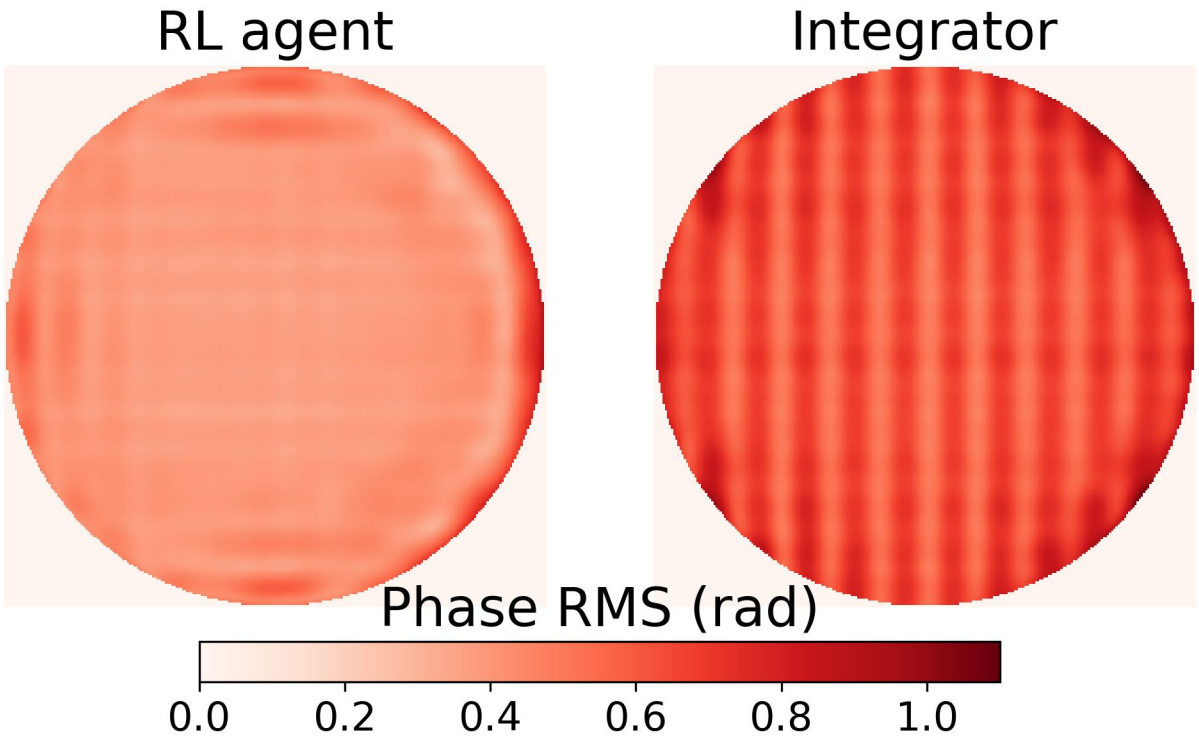
Intermezzo: Neural Networks

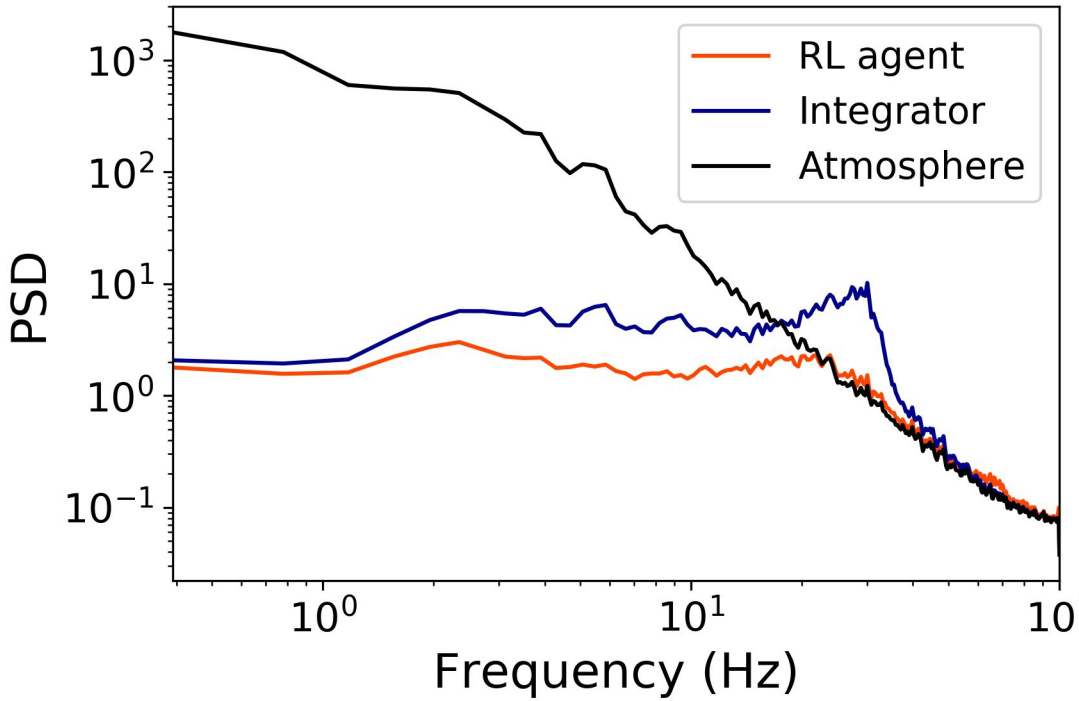
Spatio-temporal data:

Combine CNN and RNN

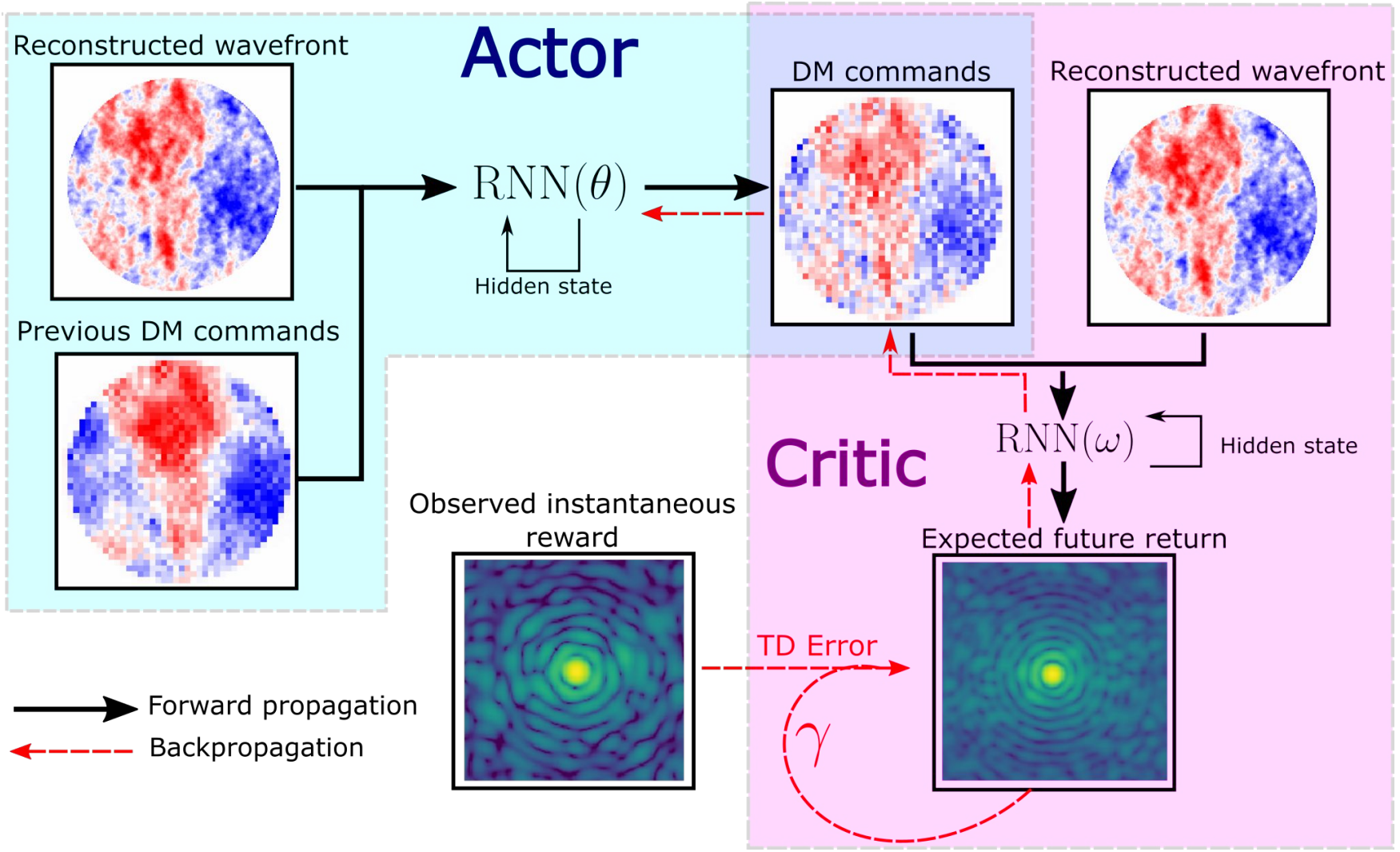


Residual power spectrum full wavefront

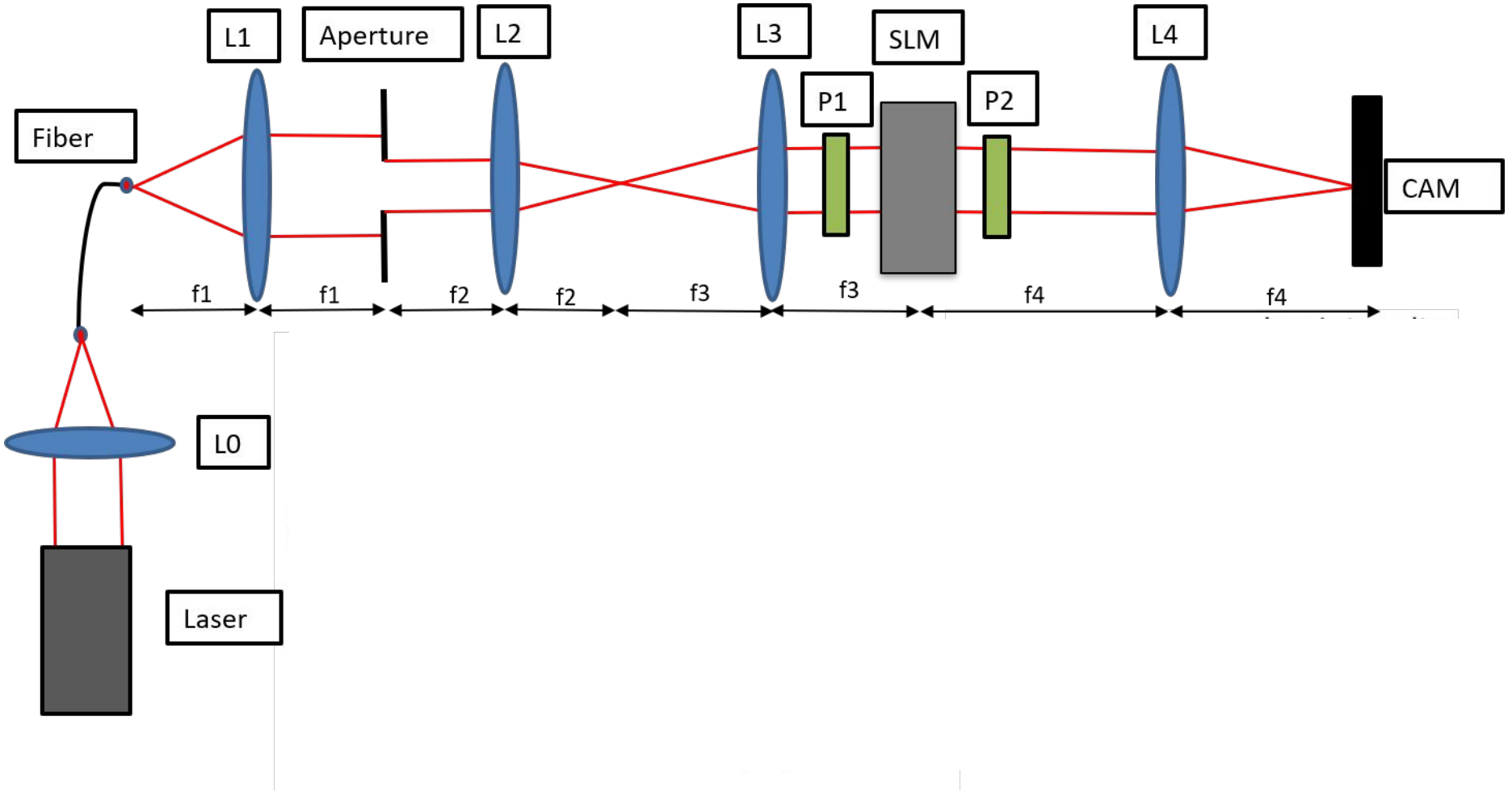




Reinforcement Learning Control (RLC)



Lab setup


L1
Aperture
L2
L3
SLM
L4
P1
P2
Fiber
CAM
f1
f1
f2
f2
f3
f3
f4
f4
L0
Laser